Techniques and Sample size

Outline

- Introduction
- definition of terms
- Types of sampling techniques
  - -probability sampling
  - -non probability sampling
- Advantage and disadvantage of sampling
- Sample size determination

Introduction

Sampling

- It is taking representative subgroup of the **reference population.**
- This is for the intension of generalization,or
- used to make estimates of the population of interest.

Key Points During Sampling

- Selecting the right people.
- Selecting the right number of the right people.
- Generally, the larger the sample size the smaller the statistical error but the greater the cost

Definition of terms

- ***Target population***

  - *popn of interest that the findings is to be generalized*

- **Source (Reference) population**.-from which the study subjects are obtained
- ***Study or Sample population***

  - the population included in the sample
- ***Sampling unit***
  - Any single unit being researched or the unit of selection in the sampling process

December 20, 2025

… Cont'd

- ***Study unit***

  - the unit in which information is collected from

- **Sampling frame**

  - the list of units from which the sample is selected

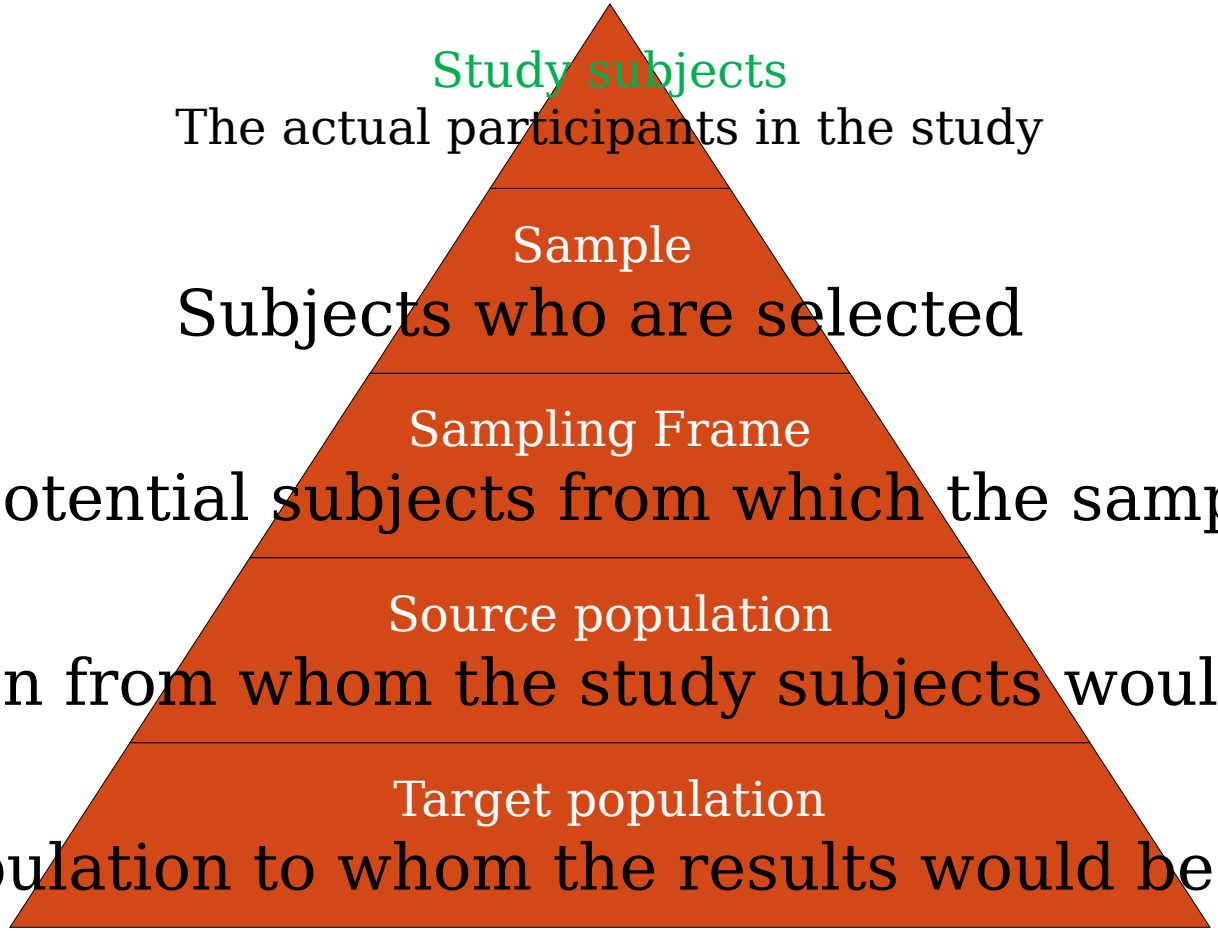
- ***Sampling fraction***

  - the ratio of the number of units in the sample to the number of units in the reference population (**n/N**).

December 20, 2025

Introduction…

The hierarchy of sampling



Reasons for Taking a Sample

- ***Practical considerations such as cost and population size***
- ***Inability of researcher to analyze large quantities of data potentially generated by a census***
- ***Samples can produce sound results if proper rules are followed for the draw***

Advantages and disadvantages of sampling

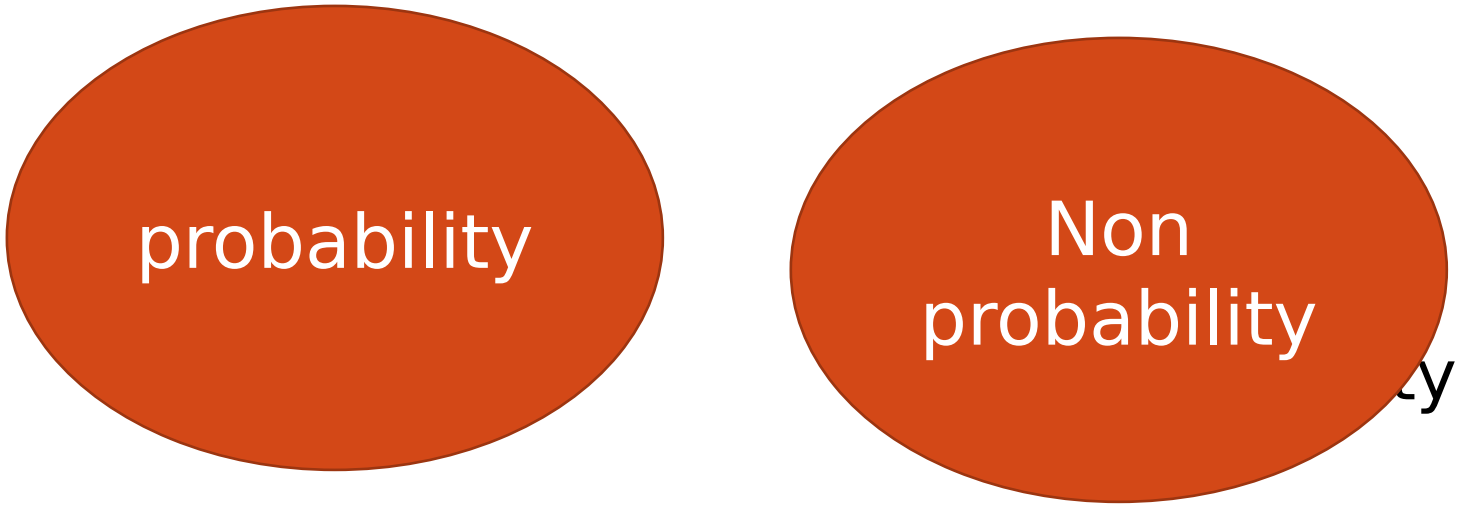
Advantages

- **Reduced cost**
- **Saves time**
- **practical**

disadvantages

- **There is sampling error**
- **May create a feeling of discrimination in the population.**

Types of sampling techniques



December 20, 2025

- ***Probability sampling:*** **ones in which members of the population have a known chance (probability) of being selected**
- ***Non-probability sampling:*** **instances in which the chances (probability) of selecting members from the population are unknown**

December 20, 2023

Types of Sampling Methods

Probability

- Simple random sampling
- Systematic random sampling
- Stratified random sampling
- Cluster sampling
- Multistage Sampling

Non probabilit

* Convenience sampling
* Judgment sampling
* Quota sampling
* Snowball sampling

A. Probability sampling Technique

* Involves random selection of a sample
* more complex
* more time-consuming and
* usually more costly than non-probability sampling

Advantages and disadvantages probability sampling...

Advantages

- **Selection is based on the principle of randomization or chance**
- **Reliable estimates can be produced**
- **Generalization can be made about the population.**

Disadvantages

- **More complex, more time-consuming and usually more costly than non-probability sampling**

Probability sampling Technique…

1.Simple Random Sampling

- Is probability sampling in which every unit has an equal nonzero chance of being selected.
- Good for **homogeneous population**

  - **SRS is drawn by the use of**
    - ✓**random number table**
    - ✓**lottery method**
    - ✓**random number generated by computer**

Cont...

- **Random number table** :It is a table of numbers constructed by a process that in any position in the table, each of the numbers 0 through 9 has probability 1/10 of occurring.
- The occurrence of any number in one part of the table is independent of the occurrence of any number in any other part of the table.

Example

- Suppose your school has 3000 students and you need to conduct a short survey on the quality of the food served in the cafeteria.
- You decide that a sample of 100 students should be sufficient for your purposes.
- In order to get your sample, you assign a number from 1 to 3000 to each student in your school.

Cont…

- To select the sample, you use a table of randomly generated numbers.
- Pick a starting point in the table (a row and column number) and look at the random numbers that appear there.
- In this case, since the data run into four digits, the random numbers would need to contain four digits as well
- Ignore all random numbers after 3000 because they do not correspond to any of the students in the school.

Table of random numbers

57172	42088	70098	11333	26902	29959	43909	49607
33883	87680	28923	15659	09839	45817	89405	70743
77950	67344	10609	87119	15859	74577	42791	75889
11607	11596	01796	24498	17009	67119	00614	49529
56149	55678	38169	47228	49931	94303	67448	31286
80719	65101	77729	83949	83358	75230	56624	27549
93809	19505	82000	79068	45552	86776	48980	56684
40950	86216	48161	17646	24164	35513	94057	51834
12182	59744	65695	83710	41125	14291	74773	66391
13382	48076	73151	48724	35670	38453	63154	58116
38629	94576	48859	75654	17152	66516	78796	73099
60728	32063	12431	23898	23683	10853	04038	75246
01881	99056	46747	08846	01331	88163	74462	14551
23094	29831	95387	23917	07421	97869	88092	72201
15243	21100	48125	05243	16181	39641	36970	99522
53501	58431	68149	25405	23463	49168	02048	31522
07698	24181	01161	01527	17046	31460	91507	16050
22921	25930	79579	43488	13211	71120	91715	49881
68127	00501	37484	99278	28751	80855	02035	10910
55309	10713	36439	65660	72554	77021	46279	22705
92034	90892	69853	06175	61221	76825	18239	47687
50612	84077	41387	54107	09190	74305	68196	75634
81415	98504	32168	17822	49946	37545	47201	85224
38461	44528	30953	08633	08049	68698	08759	45611
07556	24587	88753	71626	64864	54986	38964	83534
60557	50031	75829	05622	30237	77795	41870	26300

…Cont'd

- **Advantage**
  - **It is simple, easy to apply when the population is small.**
  - **Known and equal chance of selection**
- **Disadvantage**
  - **It is difficult to use this method for the large population in the absence of sampling frame.**
  - Requires a sampling frame.

December 20
025

2. Systematic random sampling

- **Some times called *interval sampling***
- Individuals are chosen at regular intervals(every k^{th}) from the sampling frame
- A method of probability sampling in which the defined target population is ordered and the sample is selected according to position using a skip interval

Systematic random sampling…

- **Steps in systematic random sampling**
  - **Number the units on your frame from 1 to N**
  - **Determine the sampling interval (K)**
  - **Select a number between one and K at random. This number is called *the random start* and would be the first number included in your sample**

Systematic random sampling cont…

N = 100

want n = 20

N/n = 5

select a random number from 1-5: chose 4

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Systematic sampling…

Advantages

- Easier to perform it and Require less time than SRS
- Very good when the population from which sample is to be drawn is homogeneously distributed

Disadvantage:

- ***Can not be used when a cyclic repetition is inherent (***Patterns/periodicity) ***in the sampling frame***
- ***The random start decides the chance of the rest sample units***

3. Stratified sampling

- It is appropriate when the distribution of the characteristic to be studied is heterogeneous.

1. The population is divided into groups called **strata** having similar characteristics.

- Maximum homogeneity within the group and max. heterogeneity among the groups contribute for the accuracy of the estimates.

Stratified sampling…

2. A separate sample is then taken from each stratum by random sampling

- The sample size can vary from one stratum to another.
- **Proportionate allocation-** when sample size proportional to the source population is taken from each stratum
- **Non-proportionate allocation-** the strata unequal in size and a fixed number of units is selected from each stratum

Stratified sampling...

Advantages

- representativeness of the sample is improved.
- improves the accuracy of estimation

Stratified sampling...

Disadvantages

- **problems if strata not clearly defined.**
- **Analysis is (or can be) quite complicated.**
- not useful when there are no homogeneous subgroups
- can be expensive

4. Cluster Sampling

- Reference population (homogeneous) is divided into clusters – often geographical units
- A sample of such cluster is selected
- All the units in the selected cluster are studied
- To reduce costs, researchers may choose a cluster sampling technique

…cont’d

- **Steps in cluster sampling**
  - **Divide the population into groups or clusters.**
  - **A number of clusters are selected randomly to represent the total population, and then all units within selected clusters are included in the sample.**
  - **No units from non-selected clusters are included in the sample—they are represented by those from selected clusters.**

…Cont’d

- **This differs from stratified sampling, where some units are selected from each strata**

Cluster sampling…

Advantage:

-Sampling frame of the reference population is not required (only list of clusters)

-Cost effective (efficient)

Disadvantages

- **clusters may not be representative of whole population**
- **analysis more complicated than for simple random sampling.**

5. Multistage sampling

- Used when the reference population is **large** and **widely scattered.**
- Selection is done in stages until the final sampling unit are arrived at.
  - Finally study subjects will be selected by SRS…
  - No need of sampling frame for the reference population.

Multistage sampling…

Advantage

- Cuts the cost of preparing the **sample frame**
- **Does not require a complete list of members in the target population**

Disadvantage

- sampling error is high compared with simple random sampling

B. Non Probability sampling techniques

- A type of sampling where each study unit in the population has an unknown probability of inclusion in the sample.
- The selection of subjects is **arbitrary or subjective**

When to use Non probability Sampling

- Group that represents the target population already exists.
- Difficult or impossible to obtain the list of names for sampling (**Homeless**).
- For rare population.

Non probability sampling…

Advantages

- Used when a sampling frame does not exist.
- they are quick, inexpensive and convenient(suitable).
- Good for pretests, pilot studies, In-depth interviews.
- Used when Precise representativeness is not necessary.

sampling…

Disadvantage

- No random selection (unrepresentative).
- Reliability cannot be measured.
- No way to measure the precision of the resulting sample.
- Inappropriate for generalizing findings.

Non probability sampling…

Includes

- Judgmental /Purposive
- Quota
- Convenience / haphazard
- Snow ball
- Voluntary/self selection …etc

1.Judgmental /Purposive

- The researchers choose the sample based on who they think would be appropriate for the study.
- Primarily used when there is a limited number of people that have expertise in the area being researched.
- Appropriate when the study subjects are difficult to locate.
- More efficient and economic where the sample sizes are small.
- Used where randomization is not expected to provide representative samples.

2.Quota

- The population is first segmented in to mutually exclusive sub-groups as in stratified sampling.
- Select subjects until a specific number of units/quota/ for various sub-groups has been filled.
- No rules for selecting the subjects.
- This is one of the most common forms of non-probability sampling.

3.Convenience

- Selection of subjects based on easy availability & accessibility
  - Examples :People who just happen walking
- Often used in face to face interviews

Advantage - very easy to carry out,

- Difficult to draw any meaningful conclusion.
- May not be representative

4.Snowball

- Involves a process of “chain referrals”
- Suitable for locating key informants.
- You start with one or two key informants and ask them if they know persons who know a lot about your topic of interest.
- Used when trying to interview **hard to reach groups.**

5.Volunteer/self selection

- Subjects selected are volunteers who show interest to the study.
- Common in trials demanding long duration.
- Payments for subjects some times be involved.
- Introduces strong bias/self selection bias.

Errors in sampling

1) Sampling error: Errors introduced due to the selection of a sample.

- They cannot be avoided or totally eliminated.

2) Non-sampling error (bias):

Accessibility bias

seasonality bias

Volunteer bias

Non response bias

Errors in editing and tabulation of data…etc

Sample size determination

- *How many people to study?*

Sample size determination depends on

- The objective of the study
- The study design
- Degree of precision required for generalization.
- Degree of confidence with which to conclude.

Strategies For Determining Sample Size

- Using a census for small populations

- calculate the sample size
    1. Using a computer package
    2. Using formulas to calculate a Sample Size

December 20, 2025

Using A Census For Small Populations

- Use the entire population as the sample.
- Attractive for small populations (200 or less).
- Eliminates sampling error and provides data on all the individuals in the population.

December 20, 2025

calculate the sample size

- 1. Using a computer package

  - Packages available:

    > Epinfo

    > (download free from http://www.cdc.gov/epiinfo)

  - Sample Power

  - Egret siz

  - nQuery

2. Using formulas to calculate a Sample Size

- Three criteria usually will need to be specified to determine the appropriate sample size:

✓ 1. The level of precision:

✓ 2. The level of confidence or risk, and

✓ 3. The degree of variability in the attributes being measured

December 2025

Formulas used for sample size calculation

I. Estimating a single Population Mean

$$d = Z_{\frac{\alpha}{2}} SE$$

$$= Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \sqrt{n} = Z_{\frac{\alpha}{2}} \frac{\sigma}{d}$$

$$\Rightarrow n = Z_{\frac{\alpha}{2}}^{2} \frac{\sigma^{2}}{d^{2}}$$

…cont

- Where n = sample size.

  $z_{\alpha/2}$ = Level of confidence.

  σ^2 = variability of the variable of interest.

  d=desired precision

This formula can be used in cross sectional study to estimate one group mean.

. . .

II. To estimate single population proportion

$$d = Z_{\frac{\alpha}{2}} SE$$

$$= Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}}$$

$$\Rightarrow \sqrt{n} = Z_{\frac{\alpha}{2}} \frac{\sqrt{P(1-P)}}{d}$$

$$\Rightarrow \boxed{n = Z_{\frac{\alpha}{2}}^{2} \frac{P(1-P)}{d^{2}}}$$

In order to calculate the required sample size, you need to know the following facts:

a) The reasonable estimate of the key proportion to be studied. If you cannot guess the proportion, take it as 50%. (P)
b) Amount of error that the researcher is willing to commit (margin of error)
c) Level of confidence (z)
d) The size of the population that the sample is to represent (N).
   - ✓ If it is>10,000 no need of correction formula. But
   - ✓ if the population is< 10,000 then a smaller sample size may be required and
   - ✓ use correction formula (nf=n/(1+n/N)